NeuroImage 108 (2015) 225–233

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Neural mechanisms underlying song and speech perception can be differentiated using an illusory percept

Mark Hymers [a,*], Garreth Prendergast [a,d], Can Liu [b,1], Anja Schulze [b,1], Michellie L. Young [b,1], Stephen J. Wastling [c], Gareth J. Barker [c], Rebecca E. Millman [a]

[a] York Neuroimaging Centre, University of York, York Science Park, YO10 5NY, United Kingdom
[b] Department of Psychology, University of York, YO10 5DD, United Kingdom
[c] Institute of Psychiatry, King's College London, SE5 8AF, United Kingdom
[d] Audiology and Deafness Group, School of Psychological Sciences, University of Manchester, Manchester, M13 9PL, UK

ARTICLE INFO

Article history:
Accepted 4 December 2014
Available online 13 December 2014

Keywords:
Speech
Song
Illusion
Perception
fMRI

ABSTRACT

The issue of whether human perception of speech and song recruits integrated or dissociated neural systems is contentious. This issue is difficult to address directly since these stimulus classes differ in their physical attributes. We therefore used a compelling illusion (Deutsch et al. 2011) in which acoustically identical auditory stimuli are perceived as either speech or song. Deutsch's illusion was used in a functional MRI experiment to provide a direct, within-subject investigation of the brain regions involved in the perceptual transformation from speech into song, independent of the physical characteristics of the presented stimuli. An overall differential effect resulting from the perception of song compared with that of speech was revealed in right midposterior superior temporal sulcus/right middle temporal gyrus. A left frontotemporal network, previously implicated in higher-level cognitive analyses of music and speech, was found to co-vary with a behavioural measure of the subjective vividness of the illusion, and this effect was driven by the illusory transformation. These findings provide evidence that illusory song perception is instantiated by a network of brain regions that are predominantly shared with the speech perception network.

Introduction

Perceiving language and music constitutes two of the highest level cognitive skills evident in humans. The concept that the hierarchy of syntactic structures found in language and music result in shared perceptual representations (e.g. Koelsch et al., 2002; Patel, 2003) contrasts with the idea that such stimuli are perceived using entirely disparate neural mechanisms (e.g. Peretz and Coltheart, 2003; Rogalsky et al., 2011), whilst others propose a more emergent functional architecture (Zatorre et al., 2002). Song is a well-known example of a stimulus category which evokes both linguistic and musical perception and therefore provides an avenue with which to explore the relationship between these perceptual systems.

There is currently debate regarding the extent to which the representations of melody and lyrics are integrated or segregated during the perception of song. This issue has been examined in a wide range of experiments including integration of memory for melody and lyrics of songs (Serafine, 1984; Serafine et al., 1986), neurophysiological changes resulting from semantic and harmonic incongruities in familiar music (Besson et al., 1998; Bonnel et al., 2001), fMRI repetition suppression induced by listening to unfamiliar lyrics and tunes (Sammler et al., 2010) and modulations of BOLD response to changes in words, pitch and rhythm for both spoken and sung stimuli (Merrill et al., 2012).

Existing fMRI studies have implicated an extensive network of brain regions which show larger BOLD responses to the perception of sung stimuli as compared to speech stimuli, including bilateral anterior superior temporal gyrus (STG), superior temporal sulcus (STS), middle temporal gyrus (MTG), Heschl's gyrus (HG), planum temporale (PT) and superior frontal gyrus (SFG) as well as left inferior frontal gyrus (IFG), left pre-motor cortex (PMC) and left orbitofrontal cortex (Callan et al., 2006; Schön et al., 2010).

The question of whether speech and song recruit shared or distinct neural systems remains a contentious and controversial topic which is difficult to address directly, since linguistic and musical stimuli differ in their physical attributes. Even when the same syllable is spoken or sung significant differences in the physical properties of the spoken and sung syllable are apparent, such as the minimal and maximal fundamental frequency (F0) and amplitude variation (e.g. Angenstein et al., 2012). Physical differences between spoken and sung stimuli have introduced potential low-level confounds in previous studies designed

* Corresponding author.
E-mail address: mark.hymers@ynic.york.ac.uk (M. Hymers).
[1] Student author.

http://dx.doi.org/10.1016/j.neuroimage.2014.12.010

to examine the dissociation and/or integration of speech and song perception.

Deutsch et al. (2011) demonstrated an auditory illusion in which identical auditory stimuli may be perceived as either speech or song. Deutsch's speech-to-song illusion is achieved simply through repetition of a spoken phrase. When the spoken phrase was heard for the first time, participants rated the stimulus as speech-like. Following several repetitions of the same spoken phrase, the perception of the stimulus changed and participants rated the stimulus as song-like. The perceptual transformation did not occur if the pitch of the spoken phrase was transposed, or the order of the syllables in the spoken phrase was changed during the repetition phase of the experiment. As identical stimuli can be perceived as both speech and song, Deutsch's speech-to-song illusion provides an elegant solution to controlling auditory confounds, i.e. physical differences in speech and musical stimuli.

Tierney et al. (2013) carried out an fMRI study in which they contrasted neural activity when listeners were presented with song-like and speech-like stimuli. However, rather than using identical stimuli (i.e. Deutsch's illusion in its original form), different spoken phrases were used as song- and speech-like stimuli based upon prior behavioural judgements. Using this approach, they reported BOLD changes within bilateral anterior STG, bilateral MTG, right posterior STG, left IFG and right-lateralised activity in the inferior pre-central gyrus. In contrast, in the current fMRI study, we exploited the power of Deutsch's speech-to-song illusion and employed *physically identical* stimuli that could be perceived as either speech or song. By contrasting brain regions responsive to the percept of the same stimulus as speech-like or song-like, this approach provides a direct, within-subject investigation of the integration or dissociation of neuronal activity involved in differentially perceiving speech and song. As the stimuli are physically identical in the present study, we predict that our approach should show differences in regions of higher-level auditory cortex (e.g. anterior/posterior STG, STS and MTG) as well as higher-order, heteromodal regions including left IFG and left PMC when comparing the perception of speech and illusory song.

Materials and methods

Participants

Thirty-one native English-speaking, right-handed adults gave full informed consent to participate in the study. Before taking part in the main experiment, all participants were screened for normal hearing and absence of amusia in a double-walled sound-attenuating booth. Participants who had absolute thresholds better than 20 dB HL for octave frequencies from 250 to 8000 Hz in both ears progressed to the main experiment. Four participants did not meet this requirement. Participants were also screened using a relevant subset of the Montreal Battery for the Evaluation of Amusia (MBEA: Peretz et al., 2003). One participant did not meet this requirement. As part of the MBEA, participants were asked about the number of years of formal musical training they had received. The average number of years of formal musical training was 3.3 years (range 0–16 years) in this participant group. Of the twenty-five participants who took part in the fMRI study, 15 participants had some formal musical training and 10 participants had received no formal musical training.

Twenty-six participants (mean age 22.6 years, SD 4.0 years; 8 female) were therefore entered into the main experiment. One initial pilot subject was discarded due to technical problems with data acquisition. All data from the remaining 25 participants were analysed. Participants were not paid for taking part in the experiment. The project was approved by the Research Governance Committee, York Neuroimaging Centre, University of York and conformed to the guidelines given in the Declaration of Helsinki.

Stimuli

Auditory stimuli for the main experiment were drawn from the Institute of Electrical and Electronics Engineers sentence lists (Rothauser et al., 1969). Thirty sentences were identified which contained fragments of 4–6 syllables (mean duration 2.37 s, range 1.92 to 2.83 s) — for example “in the red hot sun”. The extracted sentence fragments were used as stimuli. The experiment layout was based around 30 “trial-sets”. The layout of each of these individual trial-sets can be seen in Fig. 1. Each trial-set consisted of three pre-presentations of a stimulus, a repetition phase based around the same stimulus and three post-presentations of the same stimulus. Each trial-set used a single stimulus from the pool of 30 fragments and each participant heard each fragment in only one trial-set. The two conditions within the experiment were termed *untransformed* and *jumbled*. The difference between the two conditions occurred only during the repetition phase of the stimulus presentation — during the pre-repetition and post-repetition phases the stimuli were always presented in their original, unmodified form (see Fig. 1). In the *untransformed* condition, the repetition phase consisted of presenting the unprocessed fragment ten times, i.e. the number of repetitions shown to cause the perceptual transformation from speech to song (Deutsch et al., 2011). This was to ensure that in the post-repetition phase, the illusory transformation had already taken place. For the *jumbled* condition the Praat software (Boersma and Weenink, 2013) was used to divide each sentence fragment into individual syllables. Five-millisecond logarithmic ramps were applied to the start and end of individual syllables which were then recombined into a *jumbled* fragment as described in Deutsch et al. (2011). The repetition phase in the *jumbled* condition then consisted of the presentation of 10 sentence fragments with different syllable orderings. No perceptual transformation was predicted to occur in the *jumbled* condition. Each participant was presented with 15 trial-sets for the *jumbled* condition and 15 trial-sets for the *untransformed* condition. The experiment was performed over three scanning blocks — each of which contained 5 *jumbled* and 5 *untransformed* trial-sets. The order of the presentation of *jumbled* and *untransformed* trial-sets within the blocks was pseudo-randomised.

In order to further minimise the difference between the *untransformed* and *jumbled* conditions, the 30 stimuli were chosen from the sentence battery such that 15 pairs of stimuli approximately matched for content were derived. As an example, for the sentence fragment “in the red hot sun”, the paired fragment was “in the hot June sun”. It should be noted that the exact content of the fragments was irrelevant as only trials in which identical sentence fragments were presented were contrasted with each other in the fMRI analysis. Thus there were two sets of 15 stimuli. For each participant, one of these sets was assigned to the *untransformed* condition and the other half to the *jumbled* condition. The assignment of stimulus sets to either the *untransformed* or the *jumbled* condition was counterbalanced across subjects. This pairing counterbalancing was an extra step to minimise any potential differences between conditions.

fMRI procedure

The noise generated by MR scanners poses serious problems to researchers who wish to carry out auditory fMRI experiments (e.g. Gaab et al., 2007a, 2007b). To alleviate some of these issues, data were acquired using Interleaved Silent Steady-State Imaging (ISSS) (Schwarzbauer et al., 2006). This method of fMRI data acquisition differs from traditional sparse imaging in that even during the quiet periods, the slice-select gradient and radio-frequency excitation pulses are applied in the normal way. However, frequency-encoding, phase-encoding and data acquisition do not take place during the quiet periods. This method allows for the acquisition of multiple temporal volumes after a quiet period, without the necessity of modelling T1 saturation effects, and has been shown to be more sensitive than

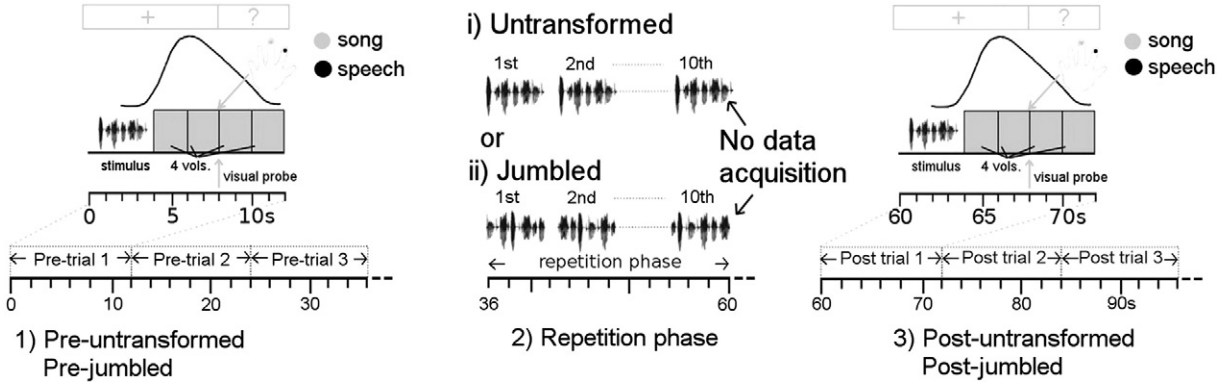

Fig. 1. Experimental layout of a single sentence fragment presentation. Each trial-set consisted of three pre-presentations, a repetition phase and three post-presentations. The pre- and post-phases involved presentation of the sentence fragment in its original form. For stimuli in the *untransformed* condition, the repetition phase involved ten repeats of the stimulus in its original form, whilst for those stimuli in the *jumbled* condition, the order of the syllables was shuffled. No fMRI data acquisition occurred during either the stimulus delivery phases of the pre- and post- presentations or during the repetition phase. Responses were visually cued 4 s into the data acquisition block after each pre- and post-presentation.

traditional sparse imaging when performing auditory fMRI experiments (Müller et al., 2011). The RMS levels of the auditory stimuli were first normalised to −25 dB FS. Participants wore earplugs and sound-attenuating headphones forming part of the fMRI-compatible auditory stimulus delivery system (MR Confon, MR Confon GmBH). The sound level of the scanner noise, not accounting for attenuation provided by earplugs and ear defenders, was 81 dB SPL during the quiet period and 98 dB SPL during the acquisition period. Stimuli were presented using Neurobehavioural Systems Presentation version 13.1 at a sound level of 98 dB SPL, not accounting for attenuation provided by earplugs only.

The experiment was divided into three fMRI runs for each participant. The order of runs was counterbalanced across participants. Each run consisted of 10 blocks, each with a 96 s duration. The layout of a block is shown in Fig. 1. Each block consisted of three pre-trials (before the repetition phase) and three post-trials (following the repetition phase). The duration of an individual trial was 12 s (giving a total of 72 s for the three pre- and three post-trials) and a repetition phase with a duration of 24 s. For each of the pre-repetition and post-repetition trials, the auditory stimulus was presented aligned to the end of a 4-s period in which no data were acquired (the quiet period). The stimulus presentation period was followed by four fMRI volume acquisitions (TR = 2 s), i.e. an 8-s duration of data acquisition during each trial. Pilot work indicated that fMRI data acquisition during the repetition phase diminished the subjective vividness of the speech-to-song illusion. Therefore no fMRI data acquisition was performed during the repetition phase, in which 10 repetitions of either the *untransformed* or *jumbled* fragment occurred. Note that the auditory stimuli during the pre- and post-trials of all conditions were the original, unscrambled fragments and the BOLD responses to these physically identical stimuli were modelled in the fMRI analyses. Within a run there were five untransformed and five jumbled blocks. All three runs, and therefore all 30 stimuli, were presented to each participant. Each participant therefore heard each stimulus either in its *untransformed* or *jumbled* context and this allocation was counterbalanced across participants. Participants were instructed to listen to the stimuli during the fMRI scans and make a response after seeing the visual cue ("?") to respond. Participants knew in advance that there were two response options, "speech" or "song" during the fMRI experiment but were not provided with explicit information about the perceptual transformation.

During the acquisition periods, whole head fMRI data (GE-EPI, TR = 2 s, TE = minimum full, flip angle = 90°) were collected using a GE Signa HDx 3 T system (General Electric, Waukesha, WI, USA). A 64 × 64 pixel matrix with a field of view of 19.2 cm was used, giving an in-plane resolution of 3 mm × 3 mm. 38 interleaved slices were collected with a slice thickness of 3 mm. A total of 245 3D volumes of data were acquired for all subjects. The volumes during the first stimulus presentation period (quiet, non-acquisition volumes) were used to allow T1 saturation to reach a steady-state.

Post-fMRI behavioural rating

As it was not possible to collect a 5-point rating in the fMRI scanner, participants were asked to provide post-scan ratings (Fig. 2). Participants were asked to listen to the same stimuli they had heard in the fMRI experiment and rate the subjective vividness of the speech-to-song illusion. Each participant was asked to rate each of the stimuli on a scale of 1 (speech-like) to 5 (song-like). The stimuli were presented using Sennheiser HD 558 headphones on a PC controlled by MATLAB (The MathWorks Inc., Natick, MA) in a quiet room.

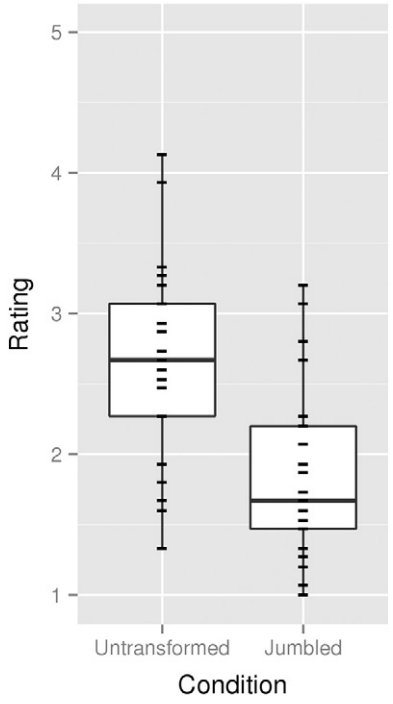

Fig. 2. Results of the post-scan behavioural experiment. Each participant was asked to rate each of the stimuli on a scale of 1 (speech-like) to 5 (song-like). There was a significant increase in rating between stimuli which had been presented as *jumbled* and *untransformed* ($t = 3.98$, $p < 0.01$, $r = 0.63$).

fMRI analysis

The fMRI data were analysed using Feat-5.98, part of FSL (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl) as well as custom scripts which implemented filtering of the temporally non-contiguous data. A separate first-level analysis was carried out for each session, for each subject. The data were motion corrected using MCFLIRT (Jenkinson et al., 2002) and brain extraction was performed using BET (Smith, 2002). The motion correction parameters were entered as regressors of no interest in the general linear model. Spatial smoothing was performed on the EPI data using a full-width half-maximum of 6 mm. Linear and quadratic trends were removed per-voxel using an in-house tool which took into account the times at which data were acquired.

Each set of three pre-repetition or post-repetition trials were modelled as separate explanatory variables (EVs). Due to the nature of the ISSS acquisition sequence, the non-contiguous temporal nature of the acquired EPI data needs to be taken into account when performing analysis. This study used a version of the analysis pipeline described in Peelle (2014), page 8. The design matrix was initially constructed to span the entire length of the experiment regardless of data acquisition. Each event entered into the design matrix consisted of the 2-second period from the onset of the auditory stimulus and was then convolved with a double gamma haemodynamic response function along with its temporal derivative (Friston et al., 1998). The design matrix was then re-sampled at the times at which fMRI acquisition occurred using a local modification to the standard FSL analysis routines (available on request from the authors). As discussed by Peelle, this avoids the necessity to adjust the degrees of freedom when assessing statistical maps at the first level. In addition, the six motion correction parameters were entered into the model and the appropriate regressor heights were recalculated for the EVs and contrasts to take into account the temporally reduced design matrix. The resulting reduced design matrix was used with FMRIB's Improved Linear Model (FILM) in order estimate beta values. Contrasts of parameter estimates were calculated by pairing each pre-repetition and post-repetition set of trials together as well as pooled estimates for each of the pre-untransformed ($\mathbf{u}_{pre}$), post-untransformed ($\mathbf{u}_{post}$), pre-jumbled ($\mathbf{j}_{pre}$) and post-jumbled ($\mathbf{j}_{post}$) conditions.

Parameter estimates were then carried through to a second-level, within-subject, fixed-effects analysis in which the mean of each condition was calculated. Finally, a third-level, between-subjects, mixed-effects analysis was performed using FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 (Beckmann et al., 2003; Woolrich et al., 2004). The primary contrast of interest was the interaction term: $(\mathbf{u}_{post} - \mathbf{u}_{pre}) - (\mathbf{j}_{post} - \mathbf{j}_{pre})$ as our primary hypothesis was that there would be differential changes in BOLD between pre-repetition and post-repetition trials in the *untransformed* compared with the *jumbled* condition. In order to disambiguate whether regions involved in the perception of the illusion either overlapped or were distinct from those involved in speech perception, a contrast involving the pre-conditions only ($\mathbf{u}_{pre}$, $\mathbf{j}_{pre}$) was performed, termed the speech-only contrast. Statistical images for all contrasts were converted to *Z* scores and corrected for multiple comparisons using a cluster-thresholding procedure (using $Z = 2.3$ and $p = 0.05$; Worsley, 2001).

Predictors for both a mean effect and a demeaned co-variate effect were included in the third-level analysis. The co-variate effect was included to reflect, for each participant, the difference between their mean ratings of the stimuli heard in the *untransformed* and *jumbled* context as measured in the post-scan behavioural experiment. This behavioural rating was included as a proxy for the subjective vividness of the illusion.

Results

Behavioural

The mean rating for stimuli heard in the *jumbled* condition (mean = 1.84, SE = 0.15) was lower than that in the *untransformed* condition (mean = 2.61, SE = 0.13). A paired *t*-test (on *jumbled* versus *untransformed* ratings) was performed and the ratings were found to differ between conditions ($t = 3.98$, $p < 0.01$, $r = 0.63$). The mean rating increase from the *jumbled* to the *untransformed* condition was 0.77 with a 95% confidence interval of 0.37 to 1.17. Consistent with a previous behavioural study of the speech-to-song illusion (Falk et al., 2014) there was no significant correlation between numbers of years of formal musical training and change in mean rating increase ($r = -0.160$; $p = 0.446$; df = 23).

fMRI

Fig. 3 illustrates the results of the interaction contrast performed at the group level $(\mathbf{u}_{post} - \mathbf{u}_{pre}) - (\mathbf{j}_{post} - \mathbf{j}_{pre})$. No statistically significant differential activations were found for the negative of the interaction term [i.e. $(\mathbf{j}_{post} - \mathbf{j}_{pre}) - (\mathbf{u}_{post} - \mathbf{u}_{pre})$], either in the mean or the co-variate analysis. The mean difference found in the interaction term (representing changes which are related to mean performance) is shown as an orange overlay. This activity localised to the right middle temporal gyrus/superior temporal sulcus (BA21/22). The co-variate analysis, based upon an individual behavioural rating (the difference between mean ratings of the *jumbled* and *untransformed* stimuli), is shown as a red overlay. A network of areas in the left frontotemporal region including the frontal pole, inferior frontal gyrus (pars opercularis), frontal orbital cortex and the temporal pole co-varied with behavioural performance (Table 1).

BOLD responses found per-participant, in the four individual conditions ($\mathbf{u}_{pre}$, $\mathbf{u}_{post}$, $\mathbf{j}_{pre}$, $\mathbf{j}_{post}$) were correlated with the participants' behavioural ratings (the covariate in the whole-brain analysis). An example of this analysis for the post-*untransformed* condition in the left frontotemporal cluster is shown to the left of Fig. 4. It should be noted that here we were interested only in which condition, or conditions, were contributing to the overall interaction; the fact that at least one of the terms contributes to a significant effect was already known. The only overall significant correlation found in all of the regions previously discussed was found in the post-*untransformed* condition ($\mathbf{u}_{pre}$: $r = 0.042$, $p = 0.840$; $\mathbf{u}_{post}$: $r = 0.595$, $p = 0.002$; $\mathbf{j}_{pre}$: $r = 0.376$, $p = 0.064$; $\mathbf{j}_{post}$: $r = 0.136$, $p = 0.517$).

To examine whether or not the resources involved in perception of the speech-to-song illusion are shared with those involved in speech perception, we compared areas significantly activated by the speech contrast with those dependent on perception of the illusion (the interaction contrast, with and without behavioural covariate) by performing a conjunction analysis. This comparison is shown in Fig. 4. The cluster-corrected activations from the speech-only contrast are shown as a blue mask. Overlap between the mean interaction contrast and the speech-only contrast are shown in green. The cluster found in the right STS/MTG in the interaction contrast consisted of 535 voxels. In a conjunction analysis with the speech-only contrast, 471 of these voxels overlapped; the remaining 64 voxels (shown in orange) were found on the inferior border of the right MTG (see Fig. 4, right-hand panel). The region (consisting of 548 voxels) which was identified in the conjunction between the co-variate analysis and the speech contrast is shown in pink (Fig. 4). The majority of the brain regions identified in the covariate analysis share neural substrates with the speech contrast, although notably regions of left frontal and fronto-orbital cortex (shown in red: 107 voxels) contribute only to the illusory percept of song, i.e. are not shared with speech perception.

To exclude the possibility that the effects were underpinned by the effect of repetition suppression during the repetition phase, an analysis of BOLD changes in the pre and post trials in the *untransformed* and *jumbled* conditions was carried out. The results of this analysis are shown in Fig. 5. In the right middle temporal regions (STS/MTG), the interaction effect was driven by decreases in BOLD relative to baseline in both conditions and the *jumbled* condition showed a greater BOLD decrease than the *untransformed* condition. In the left frontotemporal

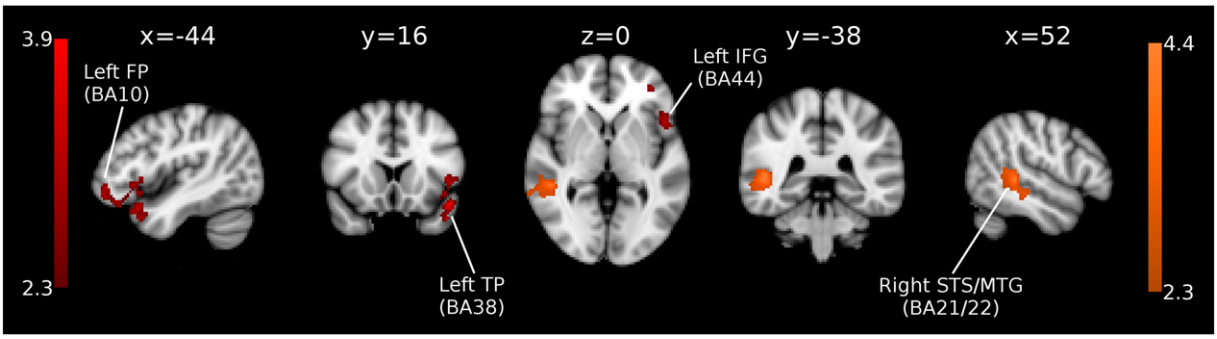

Fig. 3. Cluster-thresholded ($Z = 2.3$, $p < 0.05$), statistical maps (corrected for multiple comparisons) of the $[(\mathbf{u}_{post} - \mathbf{u}_{pre}) - (-\mathbf{j}_{post} - \mathbf{j}_{pre})]$ interaction. Areas shown in orange show a mean interaction effect whilst those in red show an effect which co-varies with the subjective vividness of the illusion. Abbreviations: STS: superior temporal sulcus; MTG: middle temporal gyrus; FP: frontal pole; IFG: inferior frontal gyrus; TP: temporal pole.

network (localised via the covariate analysis), the change between pre and post trials in the *untransformed* condition showed an increase in BOLD relative to baseline whilst the *jumbled* condition showed a decrease. If repetition suppression during the repetition phase was driving these changes, we would predict a greater BOLD decrease in the *untransformed* condition, relative to the *jumbled* condition, because the same *untransformed* fragment was presented consecutively in the *untransformed* repetition phase. However, the data are contrary to this prediction and therefore repetition suppression cannot explain the changes in BOLD response induced by listening to Deutsch's speech-to-song illusion.

Discussion

The question of whether the neural substrates of listening to speech and song reflect integrated or dissociated mechanisms was tested using Deutsch's speech-to-song illusion. Our findings provide evidence that resources recruited during speech and song perception are largely shared, notably in right midposterior STS/MTG and a left frontotemporal loop. Critically, this study differs from all previous studies on song and speech perception because physically identical, and therefore tightly controlled, stimuli were used to uncover the differential involvement of these neural systems in the perception of speech and song (cf. Tierney et al., 2013).

Shared neural substrates for speech and illusory song perception

Despite variation in the individual subjective ratings of the illusory percept, an overall differential response, i.e. the difference between perception of speech and song, was localised to regions in right midposterior STS/MTG (BA 21/22). This region has been implicated in song perception (Schön et al., 2005, 2010), melody recognition (Peretz et al., 2009), and classification of music from speech (Abrams et al., 2011). Moreover, this region has been identified as being involved in the mental imagery of song (Zatorre and Halpern, 1993; Müller et al., 2013), i.e. imaginary perception of a song when no musical stimulus is present. The identified regions in right midposterior STS/MTG overlapped completely with those areas active during the speech-only contrast. Previous findings (Zatorre and Halpern, 1993) have shown that the imagery and perception of song share neural resources and we extend this model to include the perception of speech and song stimuli, suggesting that the mechanisms in right midposterior STS/MTG reflect integrated processing.

Sammler et al. (2010) used a repetition suppression paradigm to investigate the integration and segregation of lyrics and tunes. They reported varying degrees of integration along bilateral STS/STG, with stronger integration of lyrics and tunes in more posterior auditory areas. In contrast, here we found that the effect of perceiving a stimulus as song localised to the right STS/MTG (BA 21/22). Moreover, a conjunction analysis showed that resources underlying the perception of speech and song are largely integrated. It should be noted that the tune condition in the Sammler et al. study consisted of combined variations in both melody and rhythmic content, whereas in the present study perceptual changes occurred despite the stimuli remaining acoustically identical. Despite the repeated presentation of fragments in the repetition phase of the present study, we ruled out the possibility that repetition suppression could explain the results. Sammler et al. speculated that the degree of integration/independence of lyrics and tunes may depend on the specific cognitive task required by the experiment which may have contributed to the discrepancies in degree of integration found in the two studies.

As behavioural data from individual participants showed that the subjective vividness of the illusion was variable, despite the fact that all participants were screened for normal musical ability using the

Table 1
Cluster localisation details for the interaction contrast $(\mathbf{u}_{post} - \mathbf{u}_{pre}) - (\mathbf{j}_{post} - \mathbf{j}_{pre})$. Probabilistic locations in MNI-152 space taken from the Harvard-Oxford cortical atlas.

Term	Z	Location (mm)			Area		Speech-song overlap
		x	y	z			
Mean (extent: 535 voxels)	4.40	50	−38	6	25% SMG, posterior division	BA 22	Y
	4.37	50	−38	2	30% MTG, temporo-occipital part	BA 22	Y
	3.19	52	−28	−10	27% MTG, posterior division	BA 21	Y
	3.17	56	−28	−12	20% MTG, posterior division	BA 21	Y
	3.07	58	−44	2	47% MTG, temporo-occipital part	BA 22	Y
	2.94	60	−42	18	47% SMG, posterior division	BA 22	Y
Covariate (extent: 548 voxels)	3.92	−48	18	−18	72% temporal pole (anterior)	BA 38	Y
	3.21	−44	48	−10	87% frontal pole	BA 10	N
	3.17	−48	14	0	27% IFG, pars opercularis	BA 44/45	Y
	3.13	−44	46	−6	85% frontal pole	BA 10	N
	3.11	−32	22	−16	66% frontal orbital cortex	BA 47/11	Mixed
	3.10	−50	44	−8	77% frontal pole	BA 10	Y

Abbreviations: MTG: middle temporal gyrus; SMG: supramarginal gyrus; IFG: inferior frontal gyrus.

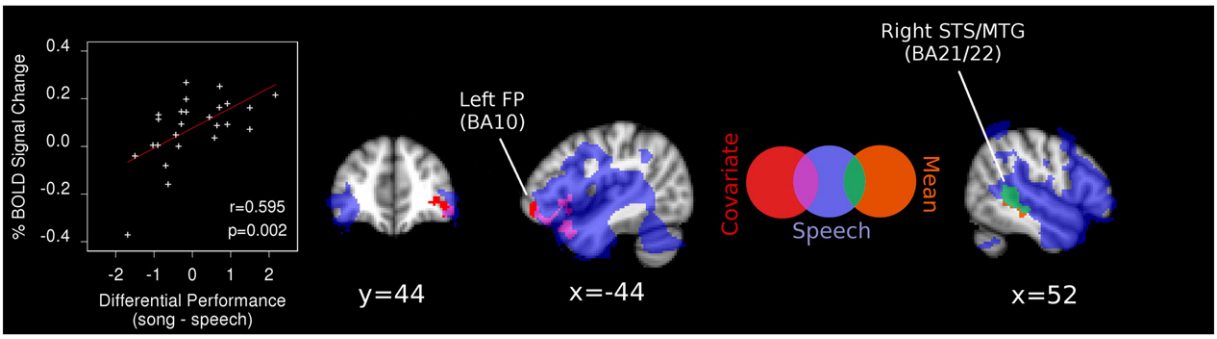

Fig. 4. Overlap and common areas of activation between the interaction term (see Fig. 3) and the mean effect of presenting speech (as defined by the speech-only contrast, shown in blue). The green area demonstrates that the majority of voxels which responded in the mean of the interaction term were also significantly active in the speech condition. The pink overlay in the left frontotemporal network shows a large area of overlap between the co-variate term and the speech-only contrast, but with the most anterior areas (shown in red) statistically significantly active in the co-variate analysis only. The far left inset on the lower row of the figure shows correlation between percent BOLD signal change and the subjective vividness of illusion on an individual participant basis. The performance rating has been de-meaned to reflect its use as a regressor for the fMRI data. This is shown in the post-untransformed case only for the left frontotemporal cluster. Other conditions are discussed in the main text. Abbreviations: STS: superior temporal sulcus; MTG: middle temporal gyrus; FP: frontal pole.

MBEA, a covariate analysis of the fMRI results was carried out to elucidate activity in brain regions that was predicted by an enhanced illusory percept. This analysis revealed that the enhanced perception of song compared with that of speech was manifest in differential activation in a network of left frontotemporal brain regions including left temporal pole (BA 38), left IFG (BA 44/45/47), left prefrontal cortex (BA 10) and left orbital cortex (BA 11). Functionally significant correlations between the BOLD change within the left frontotemporal loop and behavioural

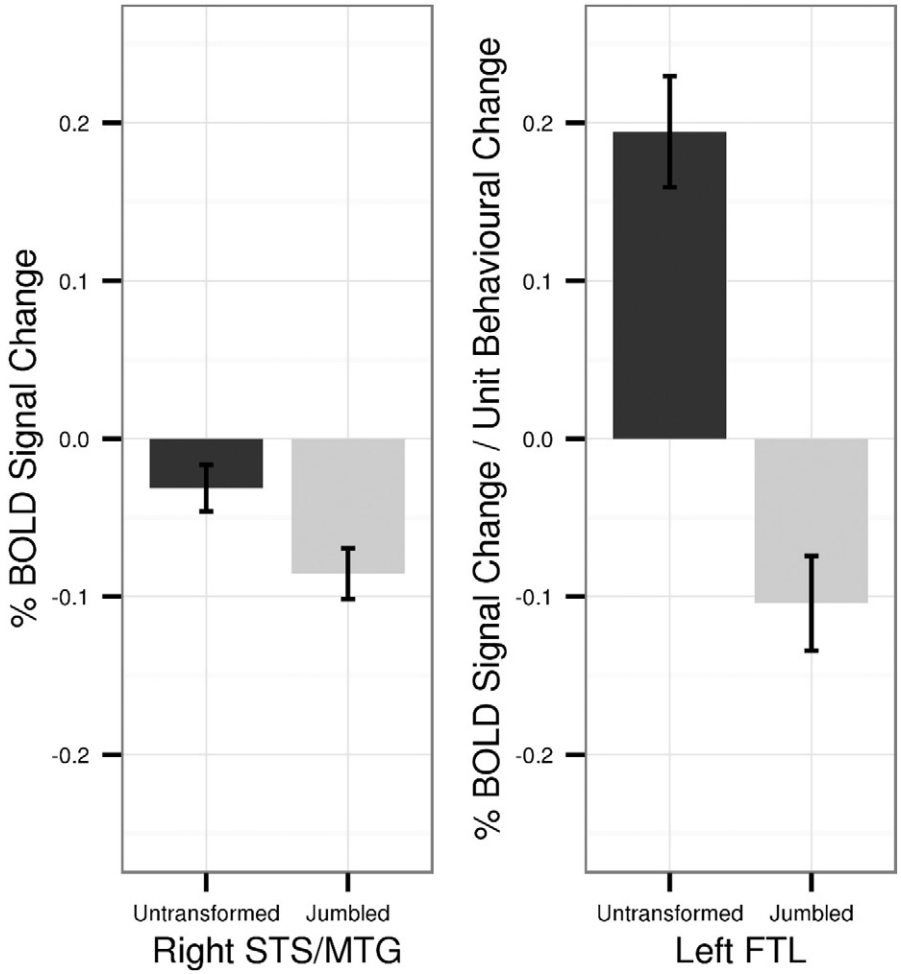

Fig. 5. BOLD changes from pre- to post-trials across *untransformed* and *jumbled* conditions. Regions were localised using the interaction contrast. Error bars show 95% confidence intervals. The left-hand panel shows the mean effect BOLD change in right posterior STS/MTG. Both conditions show a decrease from baseline, with a greater decrease in the *jumbled* compared with the *untransformed* condition. The right-hand panel shows the effect of the behavioural performance co-variate on the BOLD signal change in the left frontotemporal region. The *untransformed* condition showed an increase from baseline whilst the *jumbled* condition showed a decrease from baseline. Abbreviations: STS: superior temporal sulcus; MTG: middle temporal gyrus; FTL: frontotemporal loop.

measures of the subjective vividness of the illusion were identified only in the post-*untransformed* condition, i.e. when participants heard the stimuli as song.

The network of left frontotemporal areas identified in the current study largely overlapped with the direct, within-subject measures of speech perception (Fig. 4). Perhaps the extent of the overlap between neural systems for speech and song perception should be expected, given that song is a special musical case that requires semantic in addition to melodic analysis. However, a song stimulus is not just a linear combination of "melody + spoken lyrics" (c.f. Schön et al., 2005; Callan et al., 2006; Merrill et al., 2012) as, under normal circumstances, spoken and sung words differ in their physical properties (Angenstein et al., 2012). It therefore seems unlikely that any neural processes underlying song perception can be decomposed into entirely independent linguistic and musical cognitive processes, at least without additional integrative mechanisms.

Comparisons with previous work: listening to song vs. listening to speech

Previous work on speech and song perception (e.g. Callan et al., 2006; Schön et al., 2010) used both spoken and sung versions of the same stimuli to remove the influence of low-level articulatory properties, phonetics, syntactic structure and semantic content. However, low-level acoustical differences remain in spoken and sung versions of the same stimulus (Angenstein et al, 2012). When the perception of sung stimuli was contrasted with the perception of spoken stimuli Callan et al. (2006) found activation in areas including bilateral anterior STG, bilateral HG, bilateral PT, left premotor cortex and left orbitofrontal cortex. Schön et al. (2010) contrasted listening to the same stimulus presented as either song or speech and identified a network in bilateral STG, STS and MTG, including BA 21 and BA 22, that was lateralised towards the right hemisphere for song perception when contrasted with speech perception. In the present study the mean interaction term identified a change in the BOLD response in right midposterior STS/MTG (BA 21/22), possibly reflecting the influence of melodic processing on phonological processing (Schön et al., 2010).

Tierney et al. (2013) measured BOLD changes to speech-like and song-like speech phrases and identified an extensive network of brain regions, including bilateral anterior STG, bilateral MTG, right lateral precentral gyrus, left supramarginal gyrus, right posterior STG and left IFG. Only the brain regions identified by Tierney et al. (2013) in right posterior temporal cortex and left IFG are consistent with the regions implicated in the illusory perception of song in the present study. In contrast to the design of the present study Tierney et al. (2013) did not use the speech-to-song illusion in its original form i.e. using perceptual transformation of acoustically identical stimuli to induce the illusion. Instead Tierney et al. (2013) used different speech phrases in their "speech" and "song" conditions, which had been shown to induce the illusion in pilot testing. Tierney et al. (2013) argued that low-level acoustical differences between their "speech" and "song" conditions did not influence their results because they did not find an increase in BOLD in primary auditory cortex. However, representations of the low-level acoustical properties of sounds are not restricted to primary auditory cortex. For example, several auditory cortical areas surrounding posteromedial HG, including lateral HG, planum temporale, planum polare and superior temporal gyrus, are typically responsive to the spectrotemporal properties of sound (Griffiths et al., 2001; Patterson et al., 2002; Hall and Plack, 2009; Barker et al., 2012).

Another important difference between previous studies (Callan et al., 2006; Schön et al., 2010; Tierney et al, 2013) and the present study is that the present study used an ISSS fMRI data acquisition sequence (Schwarzbauer et al., 2006). An ISSS acquisition sequence allows for auditory presentation in the relatively quiet periods when the scanner is not acquiring data. The acoustic noise generated by fMRI scanners has several implications for auditory fMRI research, including energetic masking of auditory stimuli, reduced dynamic range in auditory cortex and increased listening effort resulting in effortful neural processing of auditory stimuli (for a recent review see Peelle, 2014).

The functional organisation of illusory song perception

According to the prevailing view of hemispheric specialisation, the left hemisphere may be more specialised for language whilst the right hemisphere is more involved in music perception. On the one hand the right-hemisphere may be more important for some aspects of musical processing such as melody perception (e.g. Samson and Zatorre, 1988; Patterson et al., 2002; Hyde et al., 2006, 2007; Albouy et al., 2013), short-term memory for pitch (e.g. Samson and Zatorre, 1991; Zatorre et al., 1992, 1994; Albouy et al., 2013) and exploration of complex acoustic environments (Teki et al., 2012). On the other hand other aspects of musical processing including pitch processing (e.g. Patterson et al., 2002), familiar melody recognition (e.g. Peretz et al., 2009) and unfamiliar song perception (Sammler et al., 2010) probably require contributions from both hemispheres.

Song is a special form of music that is more than just the sum of linguistic and musical processing (Schön et al., 2010). Therefore brain regions involved in some aspects of song perception may not necessarily be right-lateralised. Indeed, the present study found a network of brain regions in the left hemisphere, including the temporal pole, pars triangularis and orbital parts of the inferior frontal areas which co-varied with the strength of the illusory percept. These areas are typically associated with higher-level cognitive analyses of spoken language (for reviews see Binder et al., 2000; Friederici, 2011), the representations of structural regularities in music and language (Zatorre and Salimpoor, 2013), musical syntactic processing (Koelsch et al., 2004) and musical memory (Satoh et al., 2006; Platel, 1997; Platel et al., 2003; Groussard et al., 2010a,2010b). Musical semantic memory "allows us to experience a strong feeling of knowing when listening to music" (e.g. Groussard et al., 2010b). The left frontotemporal loop identified by the covariate analysis in the present study is consistent with the network for musical semantic memory (Platel, 1997; Platel et al., 2003; Groussard et al., 2010b). The overlap between speech and song in the frontotemporal loop reported here (BA 38, 44/45, and 47) is convergent with the idea that ventrolateral areas play a domain-general role in speech and song perception (Patel, 2003), although whether the computational mechanisms elucidated in these areas are identical across both modalities remains to be established.

Schön et al. (2010) suggested that left temporal and frontal brain regions may be more involved in linguistic perception whereas right temporal and frontal structures are more involved in processing the musical aspects of song. In addition they argued that anterior temporal lobe and frontal regions (BA 44/45/46/47) may be more specifically involved in the processing of complex temporal patterns. In the present study, the conjunction analysis of speech and song perception revealed largely overlapping brain regions and only some left frontal regions (BA 10/11/47) were revealed to be specific to song perception. Based on the present data, we cannot determine whether these left frontal regions (BA 10/11/47) are involved in processing complex musical patterns or some other aspects of listening to illusory song. Anterolateral frontal cortex has previously been implicated in musical semantic memory when hit-rate was included as a covariate (Groussard et al., 2010a). We interpret this as evidence that participants who perceived the illusion more strongly recruited these additional frontal regions in the left frontotemporal loop (see Figs. 3 and 4).

Overall, the results from the present study are in agreement with previous evidence that the perception of song involves both the left and right hemispheres (Callan et al., 2006; Schön et al., 2010; Sammler et al., 2010; Tierney et al., 2013). Song perception may preferentially recruit left frontotemporal regions because the linguistic aspects are an essential component of song.

Neural mechanisms underlying the speech-to-song illusion

Deutsch et al. (2011) proposed the intriguing idea that whilst listening to speech under normal conditions the neural circuitry underlying pitch salience is somewhat inhibited. This theory posits that during the repetition phase, which causes the speech-to-song illusion, the exact repetition of the speech fragment causes this circuitry to become disinhibited, thereby enhancing the salience of the perceived pitches. This leads to their prediction that activation in brain areas that respond preferentially to pitch would be enhanced. Indeed the interpretation put forward by Tierney et al. (2013) focused on mechanisms underlying pitch processing, vocalisation and auditory-motor integration. It should be noted that a model based purely on mechanisms for increased pitch saliency does not take into account that multiple auditory cues (including pitch and rhythm) contribute to the perceptual differences between speech and song (Peretz and Zatorre, 2005; Merrill et al., 2012; Falk et al, 2014). In contradistinction to the predicted pitch-based mechanism underlying the speech-to-song illusion (Deutsch et al., 2011), here we show that brain regions in right midposterior STS/MTG and a left frontotemporal loop, which are not typically implicated in low-level pitch processing, reflect the ability of participants to successfully perceive Deutsch's speech-to-song illusion.

Falk et al. (2014) hypothesised that the illusory percept of song is achieved through a mechanism of functional re-evaluation of prosodic features, supporting the idea that pitch trajectories play a major role in perceiving the speech-to-song illusion. Right temporal cortex plays a prominent role in the comprehension of prosodic information (e.g. Zatorre et al., 1992, 1994). BOLD responses specific to the evaluation of linguistic prosody occur in left lateral inferior frontal cortex (BA 44/45) (e.g. Wildgruber et al., 2004). The left inferior frontal regions (BA 44/45) implicated in the present study are therefore consistent with the idea of tracking the prosodic features of illusory song (Falk et al., 2014). As hypothesised by Falk et al. (2014) the encoded prosodic contour would then have to be interpreted as musical, possibly within the "song-specific" left frontal areas (BA 10/11/47) revealed in the present study, for the perceptual transformation to occur successfully.

Enhancing the subjective vividness of the speech-to-song illusion

The Deutsch et al. (2011) study used a phrase spoken by Diana Deutsch to successfully induce the speech-to-song illusion. Behavioural ratings of about 3.8, on a 5-point scale, for the *untransformed* condition were reported, demonstrating that the spoken phrase used in the original study resulted in a strong perceptual transformation from speech to song.

In the present study the *untransformed* condition was rated, on average, as more song-like than the jumbled control condition. However the behavioural ratings in our *untransformed* condition were not as high (mean rating of 2.61) as in the original study by Deutsch et al. (2011) (mean rating of ~3.8). One explanation for this may be that, in the present study, we used excerpts from a standard corpus of IEEE speech sentences (Rothauser et al., 1969) to test the generalisability of the speech-to-song illusion. In comparison, Tierney et al. (2013) identified phrases which resulted in the desirable perceptual transformation by an "exhaustive search" through an audiobook prior to fMRI scanning. Moreover, a recent study by Falk et al. (2014) carried out a systematic examination of the prosodic and rhythmic characteristics that are most or least likely to induce the perceptual transformation in the speech-to-song illusion. They found that tonal target stability was the most powerful cue in facilitating perceptual transformation.

Falk et al. (2014) also reported individual variation in the ability to hear the speech-to-song illusion, despite prior selection of two sentences that induced the illusion, still only 59 of 62 participants perceived the illusion and on average the perceptual transformation occurred in 65% of trials. In addition, they note that the most reliable perceptual transformations from speech to song occurred when "targeted instructions" were given to participants. The current study is in agreement with previous work (Tierney et al., 2013; Falk et al., 2014) showing that the subjective vividness of the speech-to-song illusion varies across participants and speech material used to induce the illusion. Explicitly cueing participants about the expected perceptual transformation that occurs as a result of the speech-to-song illusion, increasing the tonal target stability and providing rhythmic cues that enhance prominence contrasts of the test material may result in an improved perceptual transformation from speech to song (Falk et al., 2014).

Conclusion

Overall, our findings are in concord with the view that the perception of speech and illusory song largely share common, ventrolateral, computational substrates (Koelsch et al., 2002; Patel, 2003; Patel and Iversen, 2007; Fadiga et al., 2009) The present work demonstrates that recruitment of the left frontotemporal loop, and thereby access to brain regions crucial for higher level cognitive and semantic tasks relevant to both speech and song, relates to individual differences in subjective vividness of the speech-to-song illusion. The present findings therefore support the theory that a largely integrated network underlies the perception of speech and song.

Acknowledgments

We would like to thank Professor Andrew W. Young for helpful comments on an earlier draft of this manuscript.

Conflict of interest

SJW and GJB have received honoraria for teaching GE pulse programming during the course of this work. The remaining authors declare no competing financial interests.

References

Abrams, D.A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D.J., Menon, V., 2011. Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. Cereb. Cortex 21, 1507–1518.
Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., Tillmann, B., 2013. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. Brain 136, 1639–1661.
Angenstein, N., Scheich, H., Brechmann, A., 2012. Interaction between bottom-up and top-down effects during the processing of pitch intervals in sequences of spoken and sung syllables. Neuroimage 61, 715–722.
Barker, D., Plack, C.J., Hall, D.A., 2012. Reexamining the evidence for a pitch-sensitive region: a human fMRI study using iterated ripple noise. Cereb. Cortex 22, 745–753.
Beckmann, C.F., Jenkinson, M., Smith, S.M., 2003. General multilevel linear modeling for group analysis in FMRI. Neuroimage 20, 1052–1063.
Besson, M., Faita, F., Peretz, I., Bonnel, A.-M., Requin, J., 1998. Singing in the brain: independence of lyrics and tunes. Psychol. Sci. 9, 494–498.
Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.
Boersma, P., Weenink, D., 2013. Praat: Doing Phonetics by Computer ([Computer program]. Version 5.3.56 retrieved from http://www.praat.org/).
Bonnel, A.-M., Faita, F., Peretz, I., Besson, M., 2001. Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. Percept. Psychophys. 63, 1201–1213.
Callan, D.E., Tsytsarev, V., Hanakawa, T., Callan, A.M., Katsuhara, M., Fukuyama, H., Turner, R., 2006. Song and speech: brain regions involved with perception and covert production. Neuroimage 31, 1327–1342.
Deutsch, D., Henthorn, T., Lapidis, R., 2011. Illusory transformation from speech to song. J. Acoust. Soc. Am. 129, 2245–2252.
Fadiga, L., Craighero, L., D'Ausilio, A., 2009. Broca's area in language, action, and music. Ann. N. Y. Acad. Sci. 1169, 448–458.
Falk, S., Rathcke, T., Dalla Bella, S., 2014. When speech sounds like music. J. Exp. Psychol. 40 (4), 1491–1506.
Friederici, A.D., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91, 1357–1392.
Friston, K.J., Josephs, O., Rees, G., Turner, R., 1998. Nonlinear event-related responses in fMRI. Magn. Reson. Imaging 39, 41–52.

Gaab, N., Gabrieli, J.D.E., Glover, G.H., 2007a. Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. Hum. Brain Mapp. 28, 703–720.
Gaab, N., Gabrieli, J.D.E., Glover, G.H., 2007b. Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. Hum. Brain Mapp. 28, 721–732.
Griffiths, T.D., Uppenkamp, S., Johnsrude, I., Josephs, O., Patterson, R.D., 2001. Encoding of the temporal regularity of sound in the human brainstem. Nat. Neurosci. 4, 633–637.
Groussard, M., Rauchs, G., Landeau, B., Viader, F., Desgranges, B., Eustache, F., Platel, H., 2010a. The neural substrates of musical memory revealed by fMRI and two semantic tasks. Neuroimage 53, 1301–1309.
Groussard, M., Viader, F., Hubert, V., Landeau, B., Abbas, A., Desgranges, B., Eustache, F., Platel, H., 2010b. Musical and verbal semantic memory: two distinct neural networks? Neuroimage 49, 2764–2773.
Hall, D.A., Plack, C.J., 2009. Pitch processing sites in the human auditory brain. Cereb. Cortex 19, 576–585.
Hyde, K.L., Zatorre, R.J., Griffiths, T.D., Lerch, J.P., Peretz, I., 2006. Morphometry of the amusic brain: a two-site study. Brain 129, 2562–2570.
Hyde, K.L., Lerch, J.P., Zatorre, R.J., Griffiths, T.D., Evans, A.C., Peretz, I., 2007. Cortical thickness in congenital amusia: when less is better than more. J. Neurosci. 27, 13028–13032.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.
Koelsch, S., Gunter, T.C., v Cramon, D.Y., Zysset, S., Lohmann, G., Friederici, A.D., 2002. Bach speaks: a cortical "language-network" serves the processing of music. Neuroimage 17, 956–966.
Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A.D., 2004. Music, language and meaning: brain signatures of semantic processing. Nat. Neurosci. 7, 302–307.
Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., Friederici, A.D., 2012. Perception of words and pitch patterns in song and speech. Front. Psychol. 3, 1–12.
Müller, K., Mildner, T., Fritz, T., Lepsien, J., Schwarzbauer, C., Schroeter, M.L., Möller, H.E., 2011. Investigating brain response to music: a comparison of different fMRI acquisition schemes. Neuroimage 54, 337–343.
Müller, N., Keil, J., Obleser, J., Schulz, H., Grunwald, T., Bernays, R.-L., Huppertz, H.-J., Weisz, N., 2013. You can't stop the music: reduced auditory alpha power and coupling between auditory and memory regions facilitate the illusory perception of music during noise. Neuroimage 79, 383–393.
Patel, A.D., 2003. Language, music, syntax and the brain. Nat. Neurosci. 6, 674–681.
Patel, A.D., Iversen, J.R., 2007. The linguistic benefits of musical abilities. Trends Cogn. Sci. 11, 369–372.
Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767–776.
Peelle, J.E., 2014. Methodological challenges and solutions in auditory functional magnetic resonance imaging. Front. Neurosci. 8, 253.
Peretz, I., Coltheart, M., 2003. Modularity of music processing. Nat. Neurosci. 6, 688–691.
Peretz, I., Zatorre, R.J., 2005. Brain organization for music processing. Annu. Rev. Psychol. 56, 89–114.
Peretz, I., Champod, A.S., Hyde, K., 2003. Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. Ann. N. Y. Acad. Sci. 999, 58–75.
Peretz, I., Gosselin, N., Belin, P., Zatorre, R.J., Plailly, J., Tillmann, B., 2009. Music lexical networks: the cortical organization of music recognition. Ann. N. Y. Acad. Sci. 1169, 256–265.
Platel, H., 1997. The structural components of music perception. A functional anatomical study. Brain 120, 229–243.
Platel, H., Baron, J.-C., Desgranges, B., Bernard, F., Eustache, F., 2003. Semantic and episodic memory of music are subserved by distinct neural networks. Neuroimage 20, 244–256.
Rogalsky, C., Rong, F., Saberi, K., Hickok, G., 2011. Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. J. Neurosci. 31, 3843–3852.
Rothauser, E., Chapman, W., Guttman, N., Silbiger, H., Hecker, M., Urbanek, G., Nordby, K., Weinstock, M., 1969. IEEE recommended practice for speech quality measurements. IEEE Trans. Acoust. 17, 225–246.
Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., Samson, S., 2010. The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. J. Neurosci. 30, 3572–3578.
Samson, S., Zatorre, R.J., 1988. Melodic and harmonic discrimination following unilateral cerebral excision. Brain Cogn. 7, 348–360.
Samson, S., Zatorre, R.J., 1991. Recognition memory for text and melody of songs after unilateral temporal lobe lesion: evidence for dual encoding. J. Exp. Psychol. Learn. 17, 793–804.
Satoh, M., Takeda, K., Nagata, K., Shimosegawa, E., Kuzuhara, S., 2006. Positron-emission tomography of brain regions activated by recognition of familiar music. Am. J. Neuroradiol. 27, 1101–1106.
Schön, D., Gordon, R., Besson, M., 2005. Musical and linguistic processing in song perception. Ann. N. Y. Acad. Sci. 1060, 71–81.
Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., Besson, M., 2010. Similar cerebral networks in language, music and song perception. Neuroimage 51, 450–461.
Schwarzbauer, C., Davis, M.H., Rodd, J.M., Johnsrude, I., 2006. Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. Neuroimage 29, 774–782.
Serafine, M.L., 1984. Integration of melody and text in memory for songs. Cognition 16, 285–303.
Serafine, M.L., Davidson, J., Crowder, R.G., Repp, B.H., 1986. On the nature of melody-text integration in memory for songs. J. Mem. Lang. 25, 123–135.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.
Teki, S., Kumar, S., von Kriegstein, K., Stewart, L., Lyness, C.R., Moore, B.C., Capleton, B., Griffiths, T.D., 2012. Navigating the auditory scene: an expert role for the hippocampus. J. Neurosci. 32, 12251–12257.
Tierney, A., Dick, F., Deutsch, D., Sereno, M., 2013. Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. Cereb. Cortex 23, 249–254.
Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H., 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cereb. Cortex 14, 1384–1389.
Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Jenkinson, M., Smith, S.M., 2004. Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage 21, 1732–1747.
Worsley, K.J., 2001. Statistical analysis of activation images. Ch 14. In: Jezzard, P., Matthews, P.M., Smith, S.M. (Eds.), Functional MRI: An Introduction to Methods. Oxford University Press.
Zatorre, R.J., Halpern, A.R., 1993. Effect of unilateral temporal-lobe excision on perception and imagery of songs. Neuropsychologia 31, 221–232.
Zatorre, R.J., Salimpoor, V.N., 2013. From perception to pleasure: music and its neural substrates. Proc. Natl. Acad. Sci. U. S. A. 110, 10430–10437.
Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch processing in speech perception. Science 256, 846–849.
Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14, 1908–1919.
Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46.